Title	Measuring the coherence of normal and aphasic discourse production in Chinese using rhetorical structure theory (RST)
Author(s)	Shum, Wai-man, Waisa; 沈蕙雯
Citation	Shum, W. W. [沈蕙雯]. (2014). Measuring the coherence of normal and aphasic discourse production in Chinese using rhetorical structure theory (RST). (Thesis). University of Hong Kong, Pokfulam, Hong Kong SAR.
Issued Date	2014
URL	http://hdl.handle.net/10722/238927
Rights	The author retains all proprietary rights, (such as patent rights) and the right to use in future works.; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Measuring the Coherence of Normal and Aphasic Discourse Production in Chinese using Rhetorical Structure Theory (RST)

Shum Wai Man, Waisa

A dissertation submitted in partial fulfilment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2014

Abstract

The study investigated the difference in discourse coherence between healthy speakers and speakers with anomic aphasia using Rhetorical Structure Theory (RST). The effect of genre types on coherence and potential factors contributing to the differences were also examined. Fifteen native Cantonese participants of anomic aphasia and their control matched in age, education and gender participated. Sixty language samples were obtained using the story-telling and sequential description tasks of the Cantonese AphasiaBank protocol. Twenty naïve listeners provided subjective ratings on the coherence, completeness, correctness of order, and clarity of each speech sample. Results demonstrated that the control group showed significantly higher production fluency, total number of discourse units, and fewer errors than the aphasia group. Controls used a richer set of relations than the aphasic group, particularly those to describe settings, to express causality, and to elaborate. The aphasic group tended to omit more essential information content and was rated with significantly lower coherence and clarity than controls. The findings suggested that speakers with anomic aphasia had reduced proportion of essential information content, lower degree of elaboration, and more structural disruptions than the controls, which may have contributed to the reduced overall discourse coherence.

Introduction

Previous studies of aphasia have largely focused on the impairment in word level processing; nonetheless, language impairment can go beyond the level of lexical processing and affect the macrostructure of discourse production. Discourse production can be analyzed in terms of two main dimensions of processing: micro-linguistic and macro-linguistic abilities, responsible for intra-sentential and inter-sentential functions respectively (Glosser & Deser, 1990). The micro-linguistic level refers to the organization of phonological or orthographic patterns into morphological strings and words (lexical processing) and the formation of sentences through the organization of morphosyntactic contexts in phrases and clauses (syntactic processing) (Marino & Fabbro, 2007). In contrast, macro-linguistic dimension refers to locally the connection between sentences or utterances with the use of cohesive ties and globally the relations between all propositions in a text to integrate linguistic and conceptual features (discourse processing) (Kintsch, 1994). Cohesion refers to semantic relations among contiguous utterances, established through the use of lexical and grammatical devices such as conjunction, reference and ellipsis (Halliday & Hason, 1976). Coherence is defined by the semantic connectedness of a text at the propositional level (Van Dijk, 1980), where propositions are the minimal semantic units assigned to sentences that consist of a predicate (verb with any complements or modifiers) and a number of arguments that may have various roles. Coherence can be divided into local and global in nature. Local coherence reflects a speaker's ability to establish connection between currently processed information with the immediately preceding context (O'Brien & Albrecht, 1993). Global coherence corresponds to a speaker's ability to semantically relate remote utterances to the theme, topic, or gist of a discourse (Kintsch & Van Dijk, 1978).

Numerous studies have addressed the impairment in discourse production of people with aphasia (PWA). Some studies highlighted the microstructure of aphasic discourse in terms of lexical and syntactic processing (Saffran, Berndt, & Schwartz, 1989; Thompson, Shapiro, Li, & Schende, 1995; Vermeulen, Bastiaanse, & Van Wageningen, 1989), while others focused on the macro-linguistic abilities of PWA (Christiansen, 1995; Coelho & Flewellyn, 2003; Ulatowska, North, & Macaluso-Haynes, 1981). A multi-level approach to study both the micro- and macro-linguistic abilities of PWA was used in some other studies (Andreetta, Cantagallo, & Marini, 2012; Glosser & Deser, 1990; Wright & Capilouto, 2012). In general, studies which investigated the microstructure of discourse adopted similar quantitative measures, including measures of phonological processing (e.g. phonemic paraphasia), lexical processing (e.g. semantic paraphasia and pattern of lexical selection) and grammatical construction (e.g. syntactic complexity and errors). Nevertheless, the macrostructure of aphasic discourse was analyzed with various methods and parameters. The majority measured in terms of local and global coherence (e.g. local and global coherence errors and ratings), while others evaluated the informational content (e.g. essential proposition, lexical information unit).

Moreover, different conclusions have been drawn and led to confusing characteristics of aphasic discourse. Several reports have indicated that in spite of PWAs' impairments at the micro-linguistic level, they displayed remarkably spared macro-linguistic skills (Glosser & Deser, 1990; Ulatowska et al., 1981; Ulatowska, Weiss-Doyel, Freedman-Stern, Macaluso-Haynes, & North, 1983). Ulatowska et al. (1981; 1983) therefore suggested that these two aspects were independent to some extent. However, such dissociation was queried by researchers whose studies showed that PWA with lexical, grammatical, and phonological deficits also produced

irrelevant propositions and omitted important content. Their discourse production was characterized by the significantly lower local and global coherence ratings. Therefore, the results suggested the possibility of macro-linguistic impairment of PWA (e.g. Andreetta et al., 2012; Christiansen, 1995; Coelho & Flewellyn, 2003).

Several genre types of discourse were involved in previous studies of aphasic discourse, including story narratives (e.g. Andreetta et al., 2012; Marini, Boewe, Caltagirone, & Sergio Carlomagno, 2005; Saffran et al., 1989), personal recounts (e.g. Glosser & Deser 1990) and procedural discourse (e.g. Ulatowska et al., 1981). Different patterns of linguistic characteristics have been documented in discourse production of various genre types (Longacre, 1996), and in fact different narrative elicitation tasks might impose different cognitive and linguistic demands (Bliss & McCabe, 2006). For example, story-telling from pictures is cognitively less challenging than personal narrative and expository discourse because the latter requires organization around several ideas. Wright and Capilouto (2012) revealed significant lower coherence scores for personal recounts than story narratives. Marini et al. (2005) found significantly higher coherence for narratives elicited from describing sequential pictures compared to a single picture. The study of the influence of genre types on discourse coherence warrants further investigation owing to the limited information of possible genre effects on the micro- and macro-linguistic properties of discourse.

Regarding the aforementioned disparity in findings, Armstrong's systematic review of the literature relating to aphasic discourse (2000) suggested the differences in methodologies and theoretical bases adopted as the possible reasons for the disparity. Her review highlighted the different approaches in discourse analysis (structuralist and functionalist framework), the difference in definitions of discourse

and elicitation tasks. Although most studies claimed to have analyzed the micro- and macrostructure of narrative discourse, the definition of narrative discourse is not consistent across studies. Various elicitation tasks including picture descriptions, recounts of personal experience, story (re-)telling were used. The disparity may also be attributed to the lack of structured and systematic analysis of discourse production, which results in difficulty in the comparison of findings across studies. Moreover, although human coherence judgment is a method of assessing aphasic discourse coherence commonly used in aphasiology (e.g. Glosser & Deser 1990; Wright & Capilouto, 2012), it is arguably not sufficient to evaluate the semantic connectedness of propositions within a text just based on raters' impressions of the utterances with respect to the theme of discourse. Neither it is adequate for the investigation of the potential factors contributing to coherence. To study discourse coherence more comprehensively and objectively, systematic quantitative measures should be adopted and correlated with the subjective ratings to study the contributing factors.

In addition, the results of previous studies were limited for clinical discourse evaluation among PWA because of the small sample size in most investigations **(**e.g. 3-15 PWA in Christiansen, 1995; Glosser & Deser, 1990; Ulatowska et al., 1983). The results of case study (Coelho & Flewellyn, 2003) may not be generalized to represent the performance of PWA in general. In view of the above limitations and the complex nature of discourse, it is necessary to devise a more structured and systematic analysis to evaluate the macro-linguistic properties of discourse.

Theory driven research in discourse structure believes that discourse consists of contiguous segments bounded by semantic relations (discourse/rhetorical relations) into a global structure (Chomsky, 1975). These theories involves a variety of definitions with respect to communicative intention (Grosz & Sidner, 1986), attention

state (Walker, 1996), as well as coherence relations and cohesive devices (Hobbs, 1985; Mann & Thompson, 1988). The consensus is that discourse is structured hierarchically and not linearly. Rhetorical Structure Theory (RST) (Mann & Thompson, 1988) has been proposed in the present study as a systematic way to investigate the macro-linguistic properties of aphasic discourse. Discourse is first segmented with the use of elementary discourse units (EDUs), which refer to a clause, the minimal building blocks of the structure (Mann & Thompson, 1988). RST analyzes text organization by describing the relations that hold between units of a text in specific terms. It explains coherence by forming a hierarchical, connected structure of texts which consists of different levels.

There are similar theories concerned with relations in text such as Stack model (Grosz & Sidner, 1986) and Relevance Theory (Sperber & Wilson, 1995). Stack model integrates intentional structure with the focus of attention in discourse, while Relevance Theory presumes a relation of relevance among propositions. Despite the presence of similar theories, RST was adopted due to the following reasons. First, RST gives rise to more informative and specific relations between units of texts whereas other theories only describe a relation of dominance and relevance (i.e. one segment dominates/is relevant to the other). Second, RST produces annotation which captures the semantic features (e.g. rhetorical relations), intentional features (e.g. speaker's intention) and textual features (e.g. elements of a specific genre type) of the whole text (Carlson, Marcu, & Okurowski, 2003). By taking every piece of information into consideration, RST characterizes the types of relations between parts of the text and the depth of discourse structure, and thereby offers a comprehensive analysis (Mann & Thompson, 1988). Third, RST allows the identification of rhetorical relations in the absence of syntactic, pragmatic, or semantic marking

(Taboada, 2006). This is applicable to the discourse analysis of Chinese which is characterized by the lack of inflectional morphology (Yiu & Worrall, 1995) and frequent use of elliptical sentences (Chung, Code, & Ball, 2004). The absence of number and gender agreement between pronouns and nouns and the usual omission of topic and grammatical subjects in sentences, which are common in Chinese, therefore, do not affect the realization of semantic relations in RST. Lastly, the non-language-specific nature of the study of rhetorical relations brings about high cross-linguistic comparability. RST attributes semantic relations which hold between EDUs of a text, so the analysis is not subjected to cross-linguistic influence and therefore benefits more cross-linguistic comparison at discourse level.

In sum, this study aimed to examine differences in discourse coherence of normal speakers and PWA in connected speech using RST. It is hypothesized that PWA demonstrate significant impairment in macro-linguistic organization and hence reduced coherence. Owing to the word-finding deficit or paragrammatic errors, the omission and repetition of important propositions might be exhibited by PWA as an adaptive strategy (Christiansen 1995). This would in turn interrupt the formulation of macrostructure. Moreover, the study examined the difference in discourse coherence for the genre of narratives and procedural discourse, in light of the hypothesis that different cognitive demand is involved in different elicitation tasks. The difference in visual cues provided in the two tasks (sequential pictures in story-telling versus single picture in procedural discourse) may also influence the structure of discourse production (Fergadious, Wright, & Capilouto, 2011). We reasoned that as more visual cues were provided in the story-telling task, it would lead to a more structured discourse production. Lastly, the study investigated potential factors contributing to discourse coherence.

Method

Source of Data

The data used for the current study comes from the corpus of Cantonese aphasic discourse (Kong, Law, & Lee, 2009, November), for which data collection methods and stimuli were revised from AphasiaBank protocol (MacWhinney, Fromm, Forbes, & Holland, 2011). One-hundred and thirty six native Cantonese speakers and 76 PWA were recruited to complete nine language tasks. Language samples were collected and transcribed in the Codes for the Human Analysis of Transcripts format (CHAT: MacWhinney, 2000). Two language tasks were selected out of nine for analysis – description of making egg and ham sandwich (with one single picture presenting the ingredients for making the sandwich) and story-telling (elicited with the help of six pictures presented on the same page). The latter one is a cartoon story about a boy refusing to take an umbrella. The narrative and procedural discourse tasks were selected because we aimed to study the effect of different genre types. We chose story because of its popularity in previous studies and hence greater comparability with previous studies. Common moral story (e.g. the Tortoise and the Hare) was not chosen to avoid the effect of rote memorization because of its familiarity to speakers. The procedural discourse was selected instead of personal recount because of the greater control imposed on the content of the production.

Participants

Thirteen male and two female native Cantonese participants diagnosed of anomic aphasia according to the Cantonese version of the Western Aphasia Battery (CAB; Yiu, 1992) and their control matched in gender, age (± 5 years), and education level (± 1 year) were included. Persons with anomic aphasia were studied because most previous studies worked on discourse production of persons with anomic

aphasia. Selecting this group as the primary focus allowed greater comparability with previous studies. For the aphasic group, the age ranged between 43 and 72 (mean = 55.2 years; standard deviation = 9.70 years) and the aphasia quotients ranged from 77.1 to 99 (mean = 89.6; standard deviation = 7.09). The aphasic participants were previously recruited through the Community Rehabilitation Network, Christian Family Service Centre, as well as other non-governmental organizations and self-help groups. None of the controls had a previous history of psychiatric or neurologic illness, learning disabiilities, hearing or visual impairments.

Discourse segmentation and annotation

Every speech sample was segmented into EDUs, which were defined as syntactically a clause (Mann & Thompson, 1988). The segmentation of discourse was based on the linguistic well-formedness principles of semantic consistency and minimality (Candito & Kahane, 1998). The principles state that no elementary unit is semantically void and corresponds to more than one semantic unit. Cue phrases and prosodic contours, which are relevant types of signaling (Hirschberg & Litman, 1993), were also considered. Cue phrases are linguistic expressions that function as explicit indicators of the discourse structure. For example, “finally” and “next” are associated with the temporal relations. Prosodic contours refer to the occurrence of phrase accents and phrase-final characteristics such as pause, decrease in amplitude and phrase-final lengthening. They divide an utterance into meaningful “chunks” of information (Bolinger, 1989). Two types of units were considered as structural disruptions. The first is an incomplete EDU, which is a comprehensible clause despite inappropriate omission of words. The other is a failed EDU, which is a poorly formulated and incomprehensible clause. Embedded discourse units and comment were identified as structural expansions. The first one refers to a unit that modifies a

portion of but not the entire EDU. A comment is a subjective remark presented outside of the focus of previous units. After all EDUs were identified, adjacent text spans were linked together by rhetorical relations incrementally to form the hierarchical discourse structure. The resulting structure became wider and deeper upon continuous linkage among text spans. The depth of the structure would be the maximum number of levels attained. We used the annotation manual of the RST Treebank Corpus (Carlson & Marcu, 2001) as the reference manual and the RST Annotation Tool (version 3.4.1) as a platform for annotation. To deal with the problems of ambiguity in the definition of relations and multiple interpretations, we followed strictly the compositionality principle. It states that "rhetorical relations that hold between large textual spans can be explained in terms of rhetorical relations that hold between elementary textural units" (Marcu, 2000, p. 401). We described the relations that had been used in the current study, which were presented here and partitioned into 14 classes sharing particular types of rhetorical meaning. They include Attribution, Background, Cause, Condition, Contrast, Elaboration, Evaluation, Explanation, Joint, Manner-means, Summary, Temporal, Others and Reformulation. Three relations (correction, false-start, retracing) were added to enhance the compatibility of the framework with the current study. An example illustrating the annotated discourse structure was given in Appendix A. A description of the relations, with appropriate examples, was given in Appendix B.

Measurements of RST analysis

Each annotated sample was analyzed in terms of (1) fluency of EDU production in the first half of the sample, (2) fluency of EDU production in the second half of the sample, (3) total number of EDUs, (4) number of types of relations, (5) size of the relation set, (6) depth of the resulting discourse structure, (7) percentage of

structural expansions, (8) percentage of incomplete EDUs, (9) percentage of failed EDUs, (10) type-token ratio, (11) percentage of functional words and (12) percentage of errors (semantic, phonemic paraphasia, morphological errors, and neologisms). The measure of fluency in EDU production aimed to reflect a speaker's ability to formulate complete pieces of discourse in a definite duration. The samples were divided into halves to study the possible change in rate of production across time. The measures of the quantity and variety of rhetorical relations showed a speaker's ability to explain his way of thinking through the use of relations. The depth of structure demonstrated the degree of elaboration. When a speaker elaborated more about the same thought, the structure became deeper and more complex. The measures of structural disruption, expansion and errors were employed to investigate their effect on discourse coherence. Definitions of parameters were included in Appendix C.

Subjective ratings

Twenty students in the Division of Speech and Hearing Sciences were recruited and divided into 4 groups in a coherence rating process. Participants were asked to fill in a questionnaire of four questions for each of the 60 speech samples presented. The questionnaire comprised four questions. One question required participants to determine how well they understand the sample with 9-point scale. The other three were yes-no or multiple questions about the completeness, order and clarity of the sample. By correlating the subjective ratings with the measures from RST analysis described above, the potential factors contributing to the differences in RST analysis between groups may be revealed. Prior to the rating process, a 30-minute practice session was provided to get participants familiarized with the procedure. Items on the questionnaire and definitions of terms were explained in detail. Three trials were provided for the participants to practise. Queries and uncertainties were resolved upon

discussion during the practice session. Each group was then presented with 60 speech samples, which were presented in different random orders among groups.

Statistical analysis

Owing to the small sample size, normality was tested using Shapiro-Wilk test to determine the use of parametric or non-parametric tests. To investigate the relationship between the two genre types as well as the possible difference in performance between PWA and controls, two-way mixed analysis of variance was administered. Subsequent t-tests were carried out as post-hoc tests to study the simple main effect of genre on each of the group. An adjustment of significance level was done using Bonferroni's method due to multiple comparisons. Mann-Whiteney test and Wilcoxon rank-sum test would be done to study the group difference and effect of different elicitation tasks on each group's performance respectively if the assumption of normality was violated.

To study the difference in subjective ratings between groups, outliers were first removed from the data set with the criterion of having a z-score with an absolute value greater than 2 SD. Since only the coherence rating data was normally distributed, Mann-Whiteney test was used to compare all the measures of coherence, completeness, order, and clarity between groups. Wilcoxon signed-rank test was used to examine whether different genre types had an effect on the ratings.

To investigate the contributing factors to the coherence of discourse structure, the relationship between subjective ratings and the parameters of RST analysis was studied using the non-parametric correlation of Kendall's tau owing to the nominal nature of the subjective rating measures. The correlation coefficients between each parameter of RST analysis and each subjective rating measure were computed.

Segmentation and annotation reliability

Intra-rater reliability for EDU segmentation and annotation were determined for 10% of the samples from each group of both tasks. The samples were re-segmented and re-annotated by the same investigator. Agreements and disagreements were subjected to the following formula: (total agreements / [total agreements + total disagreements] x 100). All annotations were performed by one trained annotator, verified by a second trained annotator and reviewed by two independent judges. An agreement of the whole research group was reached for the controversial cases.

Results

Analysis of RST parameters

The Shapiro-Wilk's test was used to check the normality of the twelve RST parameters. Fluency in EDU production in the first and second half of the sample, number of types of relations and type-token ratio (TTR) were found to be normally distributed ($p > .05$). Therefore, parametric tests were implemented. The other RST parameters were not normally distributed, and therefore non-parametric tests were used. The mean values for each group on each parameter in the two elicitation tasks were reported in Appendix D.

The results of the two-way ANOVA indicated a significant main effect of group found on all parameters (see Table 1). The control group demonstrated a faster EDU production throughout the speech sample and a greater variety of relations than the aphasic group. The aphasic group showed a higher TTR than the control group. There was also a significant main effect of genre on all parameters. The story of Refused Umbrella (story-telling) yielded faster EDU productions and a larger relation set but lower TTR than the description of making sandwich (sequential description).

Interaction effect between genre type and speaker group was also significant

for fluency of the first and second half of the sample and marginally for TTR (see Fig 1). Post-hoc comparisons were conducted using independent t-tests with Bonferroni correction ($p < 0.05/8 = 0.00625$) to study the simple main effect of group in both tasks. The controls produced EDUs significantly faster in both the first ($t(28) = -3.64$, $p < .001$) and second half ($t(28) = -4.63$, $p < .001$) of the story-telling task than the aphasic group. Participants with anomic aphasia, however, described procedures of making sandwich with significantly higher TTR ($t(28) = 2.99$, $p < .00625$).

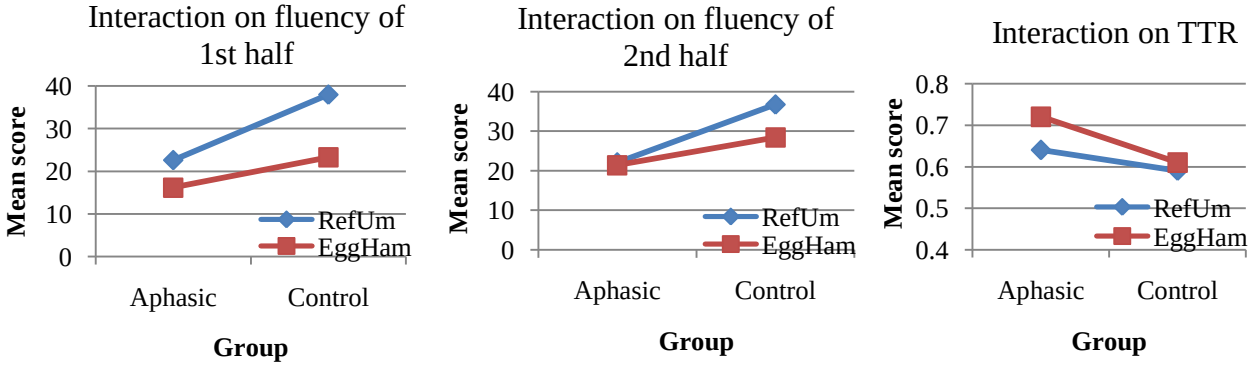
Post-hoc comparisons were conducted using dependent t-tests with Bonferroni correction ($p < 0.05/8 = 0.00625$) to investigate the simple main effect of genre on each group. Concerning the control group, the fluencyof EDU production in the first half ($t(14) = 6.54$, $p < .001$) and second half ($t(14) = 3.329$, $p < .00625$) were affected more by the effect of genre. They produced EDUs in story-telling much faster than describing procedures. The effect of genre type was stronger in the aphasic group in terms of TTR ($t(14) = -4.378$, $p < .001$), with a significantly higher TTR in sequential description than story-telling . The results showed that RST measures best differentiated between both speaker groups as well as language tasks.

Table 1.
Statistical comparisons (ANOVA) between performances of aphasic and control groups

	Two way ANOVA, F(2, 28)		
	Effect of group	Effect of genre	Interaction effect
Fluency-1st half	14.81***	28.50***	258.30*
Fluency-2nd half	18.15***	7.23*	220.19*
Types of relation	8.25**	80.33***	1.35 , p=0.60 (ns)
Type-token ratio	5.91*	7.16*	0.023, p=0.057(ms)

*Note: ns: non-significant; ms: marginally significant; *p < .05; **p ≤ .01; *** p ≤ .001*

Fig 1. *Graphs of Interaction Effects*



Regarding the results of the non-parametric tests, the aphasic group was found to produce significantly more errors in both story-telling (U = 19.00, z = -4.15, p < .001) and sequential description (U = 32.00, z = -3.77, p < .001), as well as fewer EDU in story-telling(U = 55.00, z = -2.39, p < .01) and sequential description (U = 50.50, z = -2.58, p < .01) than the control group. The control group described the story with significantly more relations (U = 44.50, z = -2.83, p < .005) than aphasic group. Wilcoxon tests showed that the story-telling elicited significantly more relations than sequential description in both control (T = 0, z = -3.41, p < .001) and aphasic groups (T = 4, z = -3.19, p < .001). The aphasic participants also produced significantly more EDUs (T = 4, z = -3.19, p < .001) and a greater depth of resultant discourse structure (T = 2, z = -3.19, p < .001) in sequential description than in story-telling task.

Analysis of subjective ratings

The significance level was set at 0.00625 (0.05/8) due to Bonferroni's adjustment . There was significant main effect of groups on coherence rating and unclarity in both story-telling and sequential description. As predicted, the control group's narratives obtained a significantly higher coherence rating and lower degree of unclarity than the aphasic group's ones in both tasks. Details of the results were shown in Table 2, 3 and 4. Mean values of subjective ratings were shown in Appendix D.

Table 2

Descriptive Statistics of the Perceptual Judgment in Story-telling Task

		Coherence rating	% Completeness	% Order	% Unclarity
Aphasic	Mean (SD)	4.60 (1.60)	42.67 (28.21)	86.67(19.52)	65.00(29.28)
	Range	1.40-6.70	5.00-95.00	25.00-100.00	10.00-100.00
Control	Mean (SD)	7.95 (0.54)	90.95 (14.19)	98.67 (2.97)	9.42 (10.17)
	Range	7.15-9.00	50.00-100.00	90.00-100.00	0.00-30.00

Table 3

Descriptive Statistics of the Perceptual Judgment in Sequential Description Task

		Coherence rating	% Completeness	% Order	% Unclarity
Aphasic	Mean (SD)	4.71 (1.74)	50.68 (34.87)	81.56 (14.45)	54.96 (28.61)
	Range	1.95-6.55	0.00-90.00	60.00-100.00	5.00-100.00
Control	Mean (SD)	7.33 (0.93)	79.21 (21.99)	94.91 (7.34)	15.21 (16.25)
	Range	6.25-9.00	25.00-100.00	73.68 -100.00	0.00-60.00

Table 4

Comparisons of speaker groups in terms of perceptual judgment by naïve listeners

Genre	Coherence rating	% Completeness	% Correct order	% Unclarity
Refused Umbrella	0.00, **p=0.000***	11.50, **p=0.000***	50.00, p=0.009 (ns)	12.00, **p=0.000***
Sandwich	8.50, **p=0.000***	51.50, p=0.010 (ns)	34.40, **p=0.001***	26.00, **p=0.000***

*Note: *Level of significance set at p=0.00625*

The effect of genre was studied with Wilcoxon's tests. No significant main effect of genre was found on any of the parameters in each group.

Analysis of relation sets

The rates of occurrence of the relation sets in each elicitation task were illustrated in graphs in Appendix E. The shapes of the graphs were similar, suggesting that the aphasic group preserved the use of relations demonstrated by the control, but did not reflect the way they used. In the story-telling tasks, the most frequently used relation type is the relations of cause for both aphasic and control groups. On the other hand, the temporal relations predominated in the sequential descriptions.

Based on visual inspection, the control group showed a tendency to use more relations of attribution, background, explanation and elaboration than the aphasic group whereas the aphasic participants produced more reformulation in both tasks. More relations of cause and condition were noted in the control's sequential description than the aphasic participants' production.

Correlation between RST parameters and subjective measures

Significant negative relationship was found in both elicitation tasks between coherence measure and two RST parameters: the percentage of errors (story-telling: $\tau = -.709$, $p < .001$; sequential description: $\tau = -.623$, $p < .001$) and the percentage of failed EDUs (story-telling: $\tau = -.389$, $p < .01$; sequential description: $\tau = -.369$, $p < .05$). In story-telling, the coherence measure was also significantly positively correlated with the fluency in EDU production in the first and second half of the sample, number of EDUs, number of types of relations, size of relation set and the depth of structure, all with the significant level of $p < .01$. The mean values of each subjective rating measure and details of correlation coefficients between RST parameters and subjective measures were listed in Appendix F.

Intra-rater reliability

Intra-rater agreement for EDU segmentation was 95.7% and that of annotation was 89.7%. The re-interpretation of the discourse content may explain the disagreement in the re-segmentation. On the other hand, disagreement in re-annotation may be accounted by the lack of clarity in the definition of relations in the RST annotation manual. It resulted in difficulty in discriminating relations with similar identification traits.

In sum, the aphasic group's discourse production was rated with a lower coherence score and degree of clarity by naïve listeners as predicted. The aphasic group produced fewer EDUs together with a reduced variety of relation types as well as increased occurrence of word-level errors. The effect of genre type was particularly evident in the elicitation of relations and rate of EDU production. Both groups demonstrated faster EDU production and a larger relation set in story-telling than sequential description. The control group produced more relations of background,

elaboration, explanation and attribution than the aphasic group, while the aphasic group produced more reformulations than the control group. The correlations between the subjective ratings and RST measures suggested that reformulations and word-level errors had a negative impact on overall coherence, whereas degree of elaboration, the quantity of EDUs and relations contributed to the maintenance of discourse coherence.

Discussion

In this study we examined the macro-linguistic property (coherence) in connected speech between 15 speakers with anomic aphasia and their healthy controls matched in gender, age and education level. The analysis was based on RST and human coherence judgment. The discourse structures of story-telling and sequential description were compared to study the effect of genre (narrative and procedural discourse). The potential factors contributing to the differences in discourse coherence between healthy speakers and speakers with anomic aphasia were also examined.

The results demonstrated a number of differences in the macro-linguistic properties of discourse production between two speaker groups. The RST measures revealed that speakers with anomic aphasia produced significantly fewer EDUs in total at a lower rate than the controls. They demonstrated reduced variety of relations than the controls in the story-telling and sequential description tasks. These findings suggested that speakers with aphasia tended to elaborate less and hence produce a less complex structure when describing both story and procedures. Since discourse structure is established through the connection between propositions with semantic relations, the reduced variety of relations may suggest that speakers with anomic aphasia could not use a wide variety of relations to illustrate their flow of thinking, such as internal reasoning (explanation, cause, reason), or motivations (e.g., purpose).

This quantitavely confirmed the observation of earlier studies which showed that the texts of speakers with mild and moderate aphasia produced oral discourse with reduced quantity and complexity (e.g. Ulatowska et al., 1981, 1983 & Korpijaakko-Huuhka & Lind, 2012).

Another main feature of the aphasic discourse was the significantly increased proportion of error production including semantic, phonemic paraphasia, morphological errors and neologisms. It was also characterized by the significantly increased production of reformulations (correction, false start and retracing) in both tasks. False starts correspond to structural errors in productions and may indicate a change of thought. The use of correction and retracing can function as a repair strategy to the errors or simply a rephrase of speech in response to a sudden change of thought or clarification. However, the excessive amounts of repair can result in disfluency in speech which would in turn hamper the clarity of speech (Tavakoli & Skehan, 2005). According to Grosz and Sidner (1986), "a discourse is coherent only when its discourse purpose is shared by all the participants, and when each utterance of the discourse contributes to achieving this purpose" (p. 202). It is hypothesized the three factors, namely word-level errors, reformulation and failed EDUs, correspond to a structural disruption. This results in unclarity that subsequently leads to difficulty in understanding the purpose of discourse, and hence reduces the discourse coherence.

It should be noted that the control group demonstrated more relations from the classes of background (background and circumstance), explanation (evidence, reason), elaboration (elaboration-additional, general-specific, process-step and set-member) and attribution than the aphasic group. The more attribution relation used might be due to a higher degree of attribution verbs to signal speech and cognitive acts. The salient relationship between the source of attribution and the reported message or

intention increased the semantic connectedness between utterances. The control group also provided more information on setting (background), causal relationships (explanation) and elaborated more. The increased elaboration provides the listener with relevant information and a complete content structure, which is more likely to be interpreted as a coherent discourse (Korpijaakko-Huuhka & Lind, 2012).

Despite the reduced elaboration in discourse production, it has been argued that the essential discourse information is relatively well-preserved by speakers with mild to moderate aphasia (Ulatowska et al., 1983; Glosser & Deser, 1990). To relate the current data with this finding, a post hoc analysis of the essential components was conducted. The details were listed in Appendix G. In the sequential description, nearly all speakers with anomic aphasia mentioned all three essential components and showed a tendency of not including optional elements of discourse. However, in story-telling, it was found that high agreement of essential components (≥80%) across speaker groups was only attained in 50% of the major proposition from orientation, complicating action and resolution. Three essential propositions from complication action and one from resolution were missing in the majority of aphasic discourse. Therefore, the findings were partially in agreement with the claim of the selectivity in reduction of content, which argued that the aphasic group only reduced elaborative material (evaluation and coda) but not the major narrative propositions (orientation, complicating action and resolution) and procedural steps (Ulatowska et al., 1983). A high proportion of information content is essentially correlated to the discourse coherence because it creates a redundancy of information such that the listener can interpret the message conveyed in discourse (Korpijaakko-Huuhka & Lind, 2012).

Effect of genre types on discourse production

Our analysis showed that both groups demonstrated faster EDU production

and a larger relation set in story-telling than in sequential description. It was hypothesized that the difference in performance across tasks may originate from the influence of the stimuli. Although stimuli were given in both tasks, they differed in terms of the amount and nature of cues. The six pictures of the story-telling task visually depicted the temporal and logical sequences of event pictures whereas only the main ingredients for making a sandwich were provided in the picture of sequential description. Therefore, the narrative task was more structured with more control on the content by stimuli (Wright & Capilouto, 2012). The visual cues in the story-telling task also provided more information on the temporal and logical relations, and may have facilitated the organization of the story structure. However, more tangentiality in sequential description may have been resulted due to the lack of a predetermined structure (Marini et al., 2005). The result confirmed our hypothesis of the effect of amount of visual cues and cognitive demand on discourse structure.

Despite the unbalanced cues across the two tasks, it was surprising to find that speakers with anomic aphasia produced significantly more EDUs and a greater depth of discourse structure in sequential description than in story-telling. It is hypothesized that the aphasic group may have been more sensitive to the effect of topic familiarity, in which previous findings suggested that familiar topics tend to elicit more elaborated production (Li, Williams, & Della Volpe, 1995). According to Britton and Tesser's cognitive model (1982), prior knowledge can be conceptualized as cognitive schema, which would be activated when dealing with a more familiar topic. Since the narrative task was more novel to the speakers than sequential description, they might show a reduced ability to formulate verbal responses and elaborate, thus reduced the depth of structure. However, as the effect was only observed in the aphasic group, further investigation is warranted to conclude on the effect of topic familiarity.

Although the effect of genre was evident in type-token ratio and the variety of semantic relations, it might not truly reflect the situation because of the small sample size. One should not overlook the effect of sample size on the TTR index values (Wright, Silverman, & Newhoff, 2003). When comparing the impact of sample size across texts of different length, the effect of sample length on TTR was found to be most significant in samples of 50-100 tokens of words than 100- to 200-token samples (Koizumi, 2012). As the sample length of both tasks in the current study was less than 100 token of words, it indicated a significant effect of sample length. When a sample is large enough, it is likely to more fully reflect the subject's active vocabulary, any further sampling of tokens can only result in a decline in TTR (Tweedie & Baayen, 1998). Given the larger means of number of words in story-telling (control: 81.4; aphasic: 61.1) than in sequential description (control: 75.9; aphasic: 36.3 words) for the aphasic group, a lower TTR in story-telling is expected. Similarly, it is reasonable to assume that a greater variety of relations can be found in a longer speech sample.

Our analysis showed that the types of semantic relations used were related to the type of discourse. It was found that temporal relations and relations of cause were mostly used by both speaker groups in sequential description and story-telling respectively. Such observation may be explained by the nature of genre type. As procedural discourse includes explanations of a sequence of events to carry out a task, it necessitates the use of temporal relations to order the events systematically. For narratives, cause and effect relationship is prominent in story-telling to connect different events to achieve semantic connectedness. This may result in the predominant use of relations of cause. As the effect of genre may predetermine the variety of relations used, the choice of task is important to the evaluation of an individual’s discourse organization and the ability to use relations to explain the way

of thinking. More types of discourse tasks should be included to obtain a more comprehensive evaluation.

Possible contributing factors to coherence

Human coherence judgment was administered to investigate the potential factors contributing to the differences in coherence of the aphasic and normal discourse. Significant differences were noted in the four measures of subjective rating between speakers with anomic aphasia and healthy speakers. Upon closer inspection the results suggested possible relationships between certain aspects of discourse processing and overall coherence. Firstly, significant difference in the rating of completeness between the speaker groups was observed in story-telling but not in sequential description. The differences may be explained by the fact that speakers with anomic aphasia preserved the essential procedural steps in sequential description but missed out several essential components in story-telling (refer to Appendix G). The more essential components the speaker preserved, the more complete the description was rated. In other words, the proportion of essential propositional content contributes to the perceptual judgment of completeness.

Since the coherence measure was found to correlate highly ($p < .001$) with the rest of the three measures (coherence completeness, order and clarity), this indicated the four parameters were likely to measure the same aspect. This may further confirm the use of subjective rating on coherence as reported in most previous studies and criticized for its validity. The high correlations among the four parameters also imply that the two factors affecting completeness and clarity, the proportion of essential propositional content and structural disruption, may also affect the overall coherence of discourse production. Moreover, the degree of elaboration and complexity of discourse structure may also play a role. This is suggested by the positive correlations

found in both tasks between coherence and the RST parameters that measured the degree of elaboration and complexity of the sample (fluency in EDU production, number of EDUs, number of relations used and depth of discourse structure).

Conclusion

Future investigations can involve a larger number of speakers with different types of aphasia and aphasia quotient (AQ) to investigate the degree of coherence across speaker groups of different AQ. Given that anomic aphasia is considered to be relatively mild, it is expected that speakers with lower AQ will show more deficits in lexical and syntactic processing and hence much lower coherence. The study of genre effect warrants the use of visual stimuli of comparable amount and nature of cues in order to discern the effects of visual cuing and genre types. Moreover, instead of conducting visual inspection on the distribution of relations, a more systematic statistical analysis should be adopted to compare the types of relations between speaker groups. One major limitation lies in the lack of inter-rater reliability measure. Therefore, more trained annotators should be involved in the replication of this study. Any disagreement aroused during annotation should be resolved upon discussion among the annotators. The inter-rater reliability should also be studied. Lastly, future research can investigate how discourse analysis can be applied clinically. The current standardized assessment on Chinese aphasic production, CAB, only uses picture description as the primary means for eliciting connected speech. Such method is not efficient because of the relatively independent events portrayed. Therefore, it may not capture the manifestations of higher-level language impairment such as maintenance of discourse coherence. Therefore, the application of RST might provide speech-language pathologists with more quantitative measures on macro-linguistic properties to devise appropriate treatment plans at discourse level for speakers with aphasia.

References

Armstrong, E. (2000). Aphasic discourse analysis: The story so far. *Aphasiology*, *14*(9), 875-892. doi:10.1080/02687030050127685

Andreetta, S., Cantagallo, A., & Marini, A. (2012). Narrative discourse in anomic aphasia. *Neuropsychologia*, *50*(8), 1787-1793. doi:10.1016/j.neuropsychologia.2012.04.003

Bliss, L. S., & McCabe, A. (2006). Comparison of discourse genres: Clinical implications. *Contemporary Issues in Communication Science and Disorders*, *33*(2), 126-137.

Bolinger, D. (1989). *Intonation and its uses: Melody in grammar and discourse*. London: Edward Arnold.

Britton, B. K., & Tesser, A. (1982). Effects of prior knowledge on use of cognitive capacity in three complex cognitive tasks. *Journal of Verbal Learning and Verbal Behavior*, *21*(4), 421-436. doi:10.1016/S0022-5371(82)90709-5

Candito, M. H., & Kahane, S. (1998). Can the TAG derivation tree represent a semantic graph? An answer in the light of Meaning-Text Theory. In *Proceedings of the Fourth Workshop on Tree-Adjoining Grammars and Related Frameworks*.

Cannito, M. P., Jarecki, J. M., & Pierce, R. S. (1986). Effects of thematic structure on syntactic comprehension in aphasia. *Brain and Language*, *27*(1), 38-49.

Carlson, L., & Marcu, D. (2001). Discourse tagging reference manual. ISI Technical Report ISI-TR-545. Retrieved from http://www.isi.edu/~marcu/discourse

Carlson, L., Marcu, D., & Okurowski., N. E. (2003). Building a discourse-tagged corpus in the framework of rhetorical structure theory. In Kuppevelt, J. V. & Smith, R. W. (Eds.), *Current and New Directions in Discourse and Dialogue* (pp.85-112). New York: Springer.

Chomsky, N. A. (1976). *Reflections on language.* London: Temple Smith.

Christiansen, J. A. (1995). Coherence violations and propositional usage in the narratives of fluent aphasics. *Brain and Language*, *51*(2), 291-317.

Chung, K. K. H., Code, C., & Ball, M. J. (2004). Lexical and non-lexical speech automatisms in aphasic Cantonese speakers. *Journal of Multilingual Communication Disorders*, *2*(1), 32-42. doi: 10.1080/1476967031000154213

Coelho, C., & Flewellyn, L. (2003). Longitudinal assessment of coherence in an adult with fluent aphasia: A follow-up study. *Aphasiology*, *17*(2), 173-182. doi: 10.1080/729255216

Fergadiotis, G., Wright, H. H., & Capilouto, G. J. (2011). Productive vocabulary across discourse types. *Aphasiology*, *25*(10), 1261-1278 doi:10.1080/02687038.2011.606974

Glosser, G., & Deser, T. (1990). Patterns of discourse production among neurological patients with fluent language disorders. *Brain and Language*, *40*(1), 67-88.

Grosz, B.J. and Sidner, C.L. (1986) Attention, Intentions, and the Structure of Discourse. *Computational Linguistics, 12*(3), 175-204.

Halliday, M. A. K., & Hasan, R. (1976*). Cohesion in English*. London: Longman.

Hirschberg, J., & Litman, D. (1993). Empirical studies on the disambiguation of cue phrases. *Computational linguistics*, *19*(3), 501-530.

Hobbs, J. (1985) *On the Coherence and Structure of Discourse.* Retrieved from http://www.isi.edu/~hobbs/ocsd.pdf

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, *49*(4), 294.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363-394.

Koizumi, R. (2012). Relationships Between Text Length and Lexical Diversity Measures: Can We Use Short Texts of Less than 100 Tokens? *Vocabulary Learning and Instruction*, *1*(1), 60-69.

Kong, A.P.H., Lee, A.S.Y., & Law, S.P. (2009, November). *The construction of a corpus of Cantonese-aphasic-discourse: a preliminary report*. Poster presented at The 2009 American Speech-Language-Hearing Association (ASHA) Convention, New Orleans, LA, USA.

Korpijaakko-Huuhka, A. M., & Lind, M. (2011). The impact of aphasia on textual coherence: Evidence from two typologically different languages. *Journal of Interactional Research in Communication Disorders*, *3*(1), 47-70.

Labov, W. (1972). *Language in the Inner City.* Philadelphia: University of Pennsylvania Press.

Li, E. C., Williams, S. E., & Della Volpe, A. (1995). The effects of topic and listener familiarity on discourse variables in procedural and narrative discourse tasks. *Journal of Communication Disorders*, *28*(1), 39-55.

Longacre, R. E. (1996). *The grammar of discourse*. New York: Plenum Press.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analysing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

MacWhinney, B., Fromm, D., Forbes, M., & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology, 25*(11), 1286-1307. doi:10.1080/02687038.2011.589893

Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text*, *8*(3), 243-281.

Marcu, D. (2000). The rhetorical parsing of unrestricted texts: A surface-based approach. *Computational Linguistics*, *26*(3), 395-448.

Marini, A., Boewe, A., Caltagirone, C., & Carlomagno, S. (2005). Age-related differences in the production of textual descriptions. *Journal of Psycholinguistic Research, 34*(5), 439-463. doi:10.1007/s10936-005-6203-z

Marini, A., & Fabbro, F. (2007). Psycholinguistic models of speech production in Bilingualism and Multilingualism. In A. Ardila & E. Ramos (Eds.), *Speech and language disorders in Bilinguals* (pp. 47-67). New York: Nova Science Publishers.

O'Brien, E. J., & Albrecht, J. E. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of experimental psychology: Learning, memory, and cognition*, *19*(5), 1061-1070. doi:10.1037/0278-7393.19.5.1061

Saffran, E. M., Berndt, R. S., & Schwartz, M. F. (1989). The quantitative analysis of agrammatic production: Procedure and data. *Brain and language*, *37*(3), 440-479.

Sperber, D., & Wilson, D. (1995).Relevance: Communication and Cognition (2nd ed.). Oxford: Blackwell.

Taboada, M. (2006). Discourse markers as signals (or not) of rhetorical relations. *Journal of Pragamatics*, *38*(4), 567-592. doi:10.1016/j.pragma.2005.09.010

Tavakoli, P., and Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239-277). Amsterdam: Benjamins.

Thompson, C. K., Shapiro, L. P., Li, L., & Schendel, L. (1995). Analysis of verbs and verb-argument structure: A method for quantification of aphasic language production. *Clinical aphasiology*, *23*(1), 121-140.

Tweedie, F. J., & Baayen, R. H. (1998). How Variable may a Constant be? Measures of Lexical Richness in Perspective. *Computers and the Humanities, 32*(5), 323-352. doi:10.1023/A:1001749303137

Ulatowska, H.K., North, A.J. & Macaluso-Haynes, S. (1981). Production of narrative and procedural discourse in aphasia. *Brain and Language, 13*(2), 345-371.

Ulatowska, H.K., Weiss-Doyell, A., Freedman-Stern, R. & Macaluso-Haynes, S. (1983). Production of narrative discourse in aphasia. *Brain and Language, 19*(2), 317-334. doi:10.1016/0093-934X(83)90074-3

Van Dijk, T. A. (1980). *Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Vermeulen, J., Bastiaanse, R., & Van Wageningen, B. (1989). Spontaneous speech in aphasia: A correlational study. *Brain and Language*, *36*(2), 252-274.

Walker, M. A. (1996). Limited attention and discourse structure. *Computational Linguistics*, *22*(2), 255-264.

Wright, H., & Capilouto, G. J. (2012). Considering a multi-level approach to understanding maintenance of global coherence in adults with aphasia. *Aphasiology*, *26*(5), 656-672. doi:10.1080/02687038.2012.676855.

Wright, H., Silverman, S., & Newhoff, M. (2003). Measures of lexical diversity in aphasia. *Aphasiology*, *17*(5), 443-452. doi:10.1044/1058-0360(2013/12-0083)

Yiu, E. M. L. (1992). Linguistic assessment of Chinese-speaking aphasics: Development of a Cantonese aphasia battery. *Journal of Neurolinguistic*, *7*(4), 379-424. doi:10.1016/0911-6044(92)90025-R

Yiu, E. M. L., & Worrall, L. (1996). Patterns of grammatical disruption in Cantonese aphasic subjects. *Asia Pacific Journal of Speech, Language and Hearing*, *1*(2), 105-126. doi:10.1179/136132896805577531.

Acknowledgments

I would like to express my deepest gratitude to my supervisors, Dr Sam-Po Law and Dr Anthony Pak-Hin Kong, for their generous support and guidance throughout the course of this dissertation. I would also like to show my greatest appreciation to Ms. Anastasia Linnik whose meticulous comments were an enormous help to me. The writing of this dissertation has been one of the greatest challenges I have ever faced throughout the pursuit of my undergraduate study. Their insightful comments and suggestions have given me concrete support in accomplishing this thesis. Without their persistent help, it would not have been possible. I also want to thank Ms. Christy Lai and Ms. Winsy Wong for their assistance in the data collection.

Most importantly, I would like to extend my sincere thanks to my family, especially my mum for her patience and well-rounded care during this tough period. Special thank to my friends who have provided me encouragement and support.

Appendix A

Example of annotated discourse structure

(1) The elementary discourse units (EDUs) are determined first.

(2) Adjacent EDUs are linked together by rhetorical relations incrementally to form the hierarchical discourse structure.

(3) Nuclearity is determined when assigning relations. A mononuclear relation contains a nucleus (important and indispensible information) and a satellite (can be deleted), e.g. text spans [1] (nucleus) and [2-3] (satellite). A multinuclear relation contains two or more units of equal importance, e.g. text spans [1-4] and [5-6]. The depth of the resulting discourse structure is 6 levels (text span [1-6]).

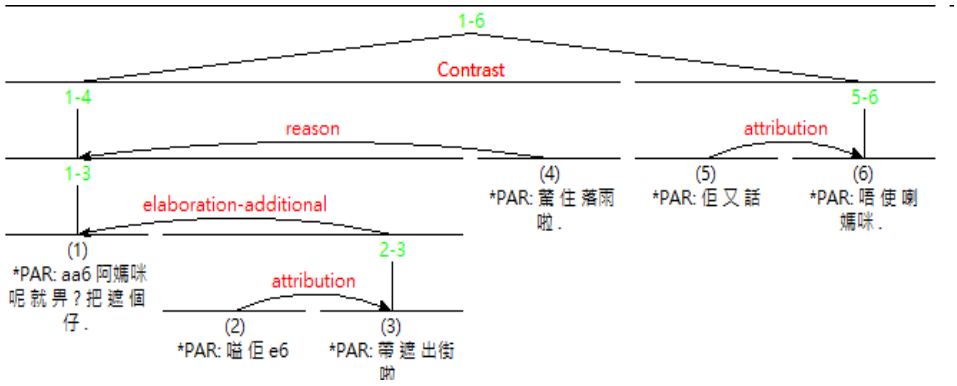


Table A1

Translation of EDUs

(1): Mum gave an umbrella to son	(2): told him	(3): bring an umbrella out
(4): feared raining	(5): then he said	(6): no need, mum

Note: Chinese is lack of inflectional morphology. It is reflected in the English translation for better understanding of the EDUs.

Table A2

Examples of Structural Disruption and Expansion (parts in bracket)

Incomplete EDU: 今日開學日，有兩個…一個母親，一個女。	Today, the first day of school, [there are two (persons)], a mother and a daughter.
Failed EDU: 佢唔攞，於是乎就後尾佢經…無攞遮啦。	He didn't bring. [Then afterwards he passes…] didn't bring an umbrella.
Embedded discourse unit: 有一個細路仔一個女人，應該係佢媽媽，嗌佢攞翻把遮。	There is one child, one woman, [should be his mum], told him to take the umbrella.
Comment: 整火腿蛋三文治要預備三種材料，[其實我唔鐘意食三文治]。	To make an egg and ham sandwich you need to prepare three ingredients, [actually I don't like sandwich.]

Appendix B

Description of relation and examples

Based on Carlson and Marcu's classification (2001), all the relations used in the annotation were partitioned into the following classes.

Table B1
Partition of Relations

Class	Relation	Class	Relation
Attribution	Attribution	Joint	Joint, List
Background	Background, Circumstance	Manner-means	Manner, Mean
Condition	Condition	Reformulation	Correction, false-start, retracing
Contrast	Contrast	Relations of Cause	Consequence, Purpose
Elaboration	Elaboration-additional, general-specific, process-step, set-member	Temporal	Temporal-after, Temporal-same-time, Sequence
Evaluation	Interpretation, Comment	Summary	Summary, Restatement
Explanation	Evidence, Reason	Others	Same-Unit, Rhetorical-question

Table B2
Definitions and Examples of Relations

Type	Example
Attribution: The satellite is the source of a reported message and the nucleus is the content of it.	[媽媽嗌佢] *[帶把遮啦]* [Mum told him] *[bring an umbrella]*
Background: The satellite establishes the context with respect to which the nucleus is to be interpreted. Unlike the 'circumstance' relation, the context is not always specified clearly.	*[有一日朝頭早]* [媽咪叫小朋友帶遮] *[There is a morning one day]* [mum told the child to bring an umbrella]
Circumstance: Being a stronger relation than a 'background' relation, the events in nucleus and satellite are co-temporal though.	*[行到半路]* [突然落起大雨喎] *[Walking halfway]* [it started to rain suddenly]
Condition: The truth of the proposition associated with the nucleus is a consequence of the fulfillment of the satellite.	*[如果你想好食啲]* [最好搽牛油] *[If you want better taste]* [better put on butter]
Contrast: Two or more nuclei come in contrast with each other along some dimension.	[媽媽叫佢帶遮] [佢唔帶] [Mum told him to bring an umbrella] [he refused to bring]

Note. *Words in italics*: *satellite*; words in normal font: nucleus

Type	Example	Type	Example
Elaboration-set-member: The nucleus introduces a finite set. The satellite elaborates on the list of members	[火腿兩種食法] *[一係切絲][一係成塊放落去麵包]* [There are two ways to eat the ham] *[shred the ham] [put the whole piece on bread]*	**Elaboration-process-step**: The nucleus introduces an event (a process). The satellite lists the steps involved.	[整雞蛋三文治] *[首先要煎熟蛋同火腰] [跟著攞係面包]* [To make egg and ham sandwich] *[first shallow-fry an egg and ham][then put on the bread]*
Interpretation: It can be multi- or mononuclear. One side of the relation gives a different perspective (an explanation or understanding) on the situation in the other side.	[小朋友話我唔帶遮] *[始終佢都係唔帶遮]* [The child said I wouldn't bring an umbrella] *[still the child doesn't bring the umbrella]*	**Comment:** The satellite constitutes a subjective remark on the nucleus.	[整火腿蛋治] *[哩個反而熟悉啲]* [To make egg and ham sandwich] *[in fact this one is more familiar]*
Evidence: The satellite provides justification for the situation in nucleus.	[天氣咁好] *[唔會落雨]* [The weather is nice] *[it won't rain]*	**Reason:** The nucleus must be an action carried out by an agent on purpose. The satellite is the reason for nucleus.	[我唔帶遮啦] *[天氣咁好]* [I won't bring an umbrella] *[the weather is nice]*
Joint: It is a pseudo-relation and unspecific in nature. It connects two segments which are not bound by a specific relation but cannot be split apart.	[小朋友攞翻把遮] [繼續返學] [The child took back the umbrella] [continued to walk to school]	**List:** It is a multinuclear relation whose elements can be listed but not in a comparison or contrast.	[第一樣材料係一隻蛋] [仲要一塊火腿] [The first ingredient is an egg] [and need a piece of ham]
Manner: The satellite explains the way in which something is done.	*[媽媽好嬲]* [話個小朋友] *[Mum very angry]* [scolded the child] (Angry is a stative verb)	**Mean:** The satellite specifies a method, instrument or mechanism for accomplishing some goal.	*[用筷子]* [打勻隻蛋] *[Use chopsticks]* [whisk the egg]
Correction: The satellite corrects the information in nucleus	[攞翻件雨褸] *[攞翻把雨傘]* [Take the raincoat] *[Take an umbrella*]	**False-start:** The satellite consists of words/phrases that are cut off mid-utterance	*[個男仔係...]* [就落雨啦] *[The boy is…]* [It rains]

Note. *Words in italics*: *satellite*; words in normal font: nucleus

Type	Example	Type	Example
Retracing: The satellite traces back to a previous utterance (nucleus) and maintains a similar idea	[媽咪見到] [有少少嬲] *[見到佢濕晒]* [Mum saw] [was a bit angry] *[saw him drenched]*	**Consequence:** The situation in one span is a consequence of another. It suggests a more indirect linkage between the events, without an explicit intention. It can be multi- or mononuclear.	[真係落雨喎] *[個小朋友濕晒]* [It really rained] *[the child was drenched]*
Purpose: The situation in the satellite is yet to be achieved by accomplishing the situation in the nucleus.	[走翻返去] *[攞翻把遮]* [Went back] *[take back an umbrella]*	**Summary:** The satellite summarizes the information presented in a nucleus.	(製作腿蛋治的步驟) *[哩個就係腿蛋治]* (Procedures in making sandwich *[This is an egg and ham sandwich]*
Restatement: The satellite reiterates the information presented in nucleus, typically with slightly different wording.	[大過都得] *[大過塊面包都可以]* [Bigger is ok] *[bigger than a piece of bread is fine]*	**Temporal-after:** The situation in the nucleus occurs after the situation in the satellite	*[煎完之後]* [就放落麵包度] *[After shallow-frying it]* [put it on the bread]
Temporal-same-time: The situations in the spans occur at approximately the same time. It can be multi- or mononuclear.	[一路行] [一路真係落雨] [walk and walk] [it really rained at the same time]	**Sequence:** Similar to 'temporal-after' relation, it refers to a set of spans (nuclei) occur in temporal order.	[首先煎隻蛋] [跟著煎火] [最後放喺麵包] [First shallow-fry the egg] [then shallow-fry the ham] [lastly put them on bread]
Same-Unit: It is a pseudo-relation that links two discontinuous text fragments which belong to a single EDU, but broken up by an embedded unit.	[一個女人] [應該係佢媽媽] [嗌佢] [A woman] [should be his mother] (embedded unit) [told him]	**Rhetorical-question:** The satellite poses a question on the nucleus. The intention was to raise an issue for the listener to consider.	[唔聽我講啦] [*點解唔帶遮*] [Not listen to what I said] [*Why didn't bring an umbrella?]*

Note. *Words in italics*: *satellite*; words in normal font: nucleus

Appendix C

Measurements of RST analysis

I. *Fluency of EDU production in the first half of the sample*: It is divide by total number of EDUs produced in the first half of the recording by the time elapsed in minutes.

II. *Fluency of EDU production in the second half of the sample*: It is divided by total number of EDUs produced in the second half of the recording by .

III. *Total number of EDUs*: It refers to the total number of EDUs produced, including semantic and phonemic paraphasiaand words in all sorts of errors.

IV. *Number of types of relation*: It refers to the total number of types of rhetorical relations used in the oral sample.

V. *Size of the relation set*: It refers to the total number of rhetorical relations used.

VI. *Depth of the resulting discourse structure*: It is equal to the maximum number of levels attained in the discourse structure.

VII. *Percentage of structural expansions*: It is divided by the number of embedded discourse unit and comment by total number of EDUs.

VIII. *Percentage of incomplete EDUs*: It is divided by the total number of incomplete EDUs by total number of EDUs.

IX. *Percentage of failed EDUs*: It is divided by total number of failed EDUs by total number of EDUs.

X. *Type-token ratio*: This is the ratio of number of different words to total number of words, excluding repetition, retracing , self-correction and false-start.

XI. *Percentage of functional words*: It is the proportion of all closed-class words with respect to total number of words.

XII. *Percentage of errors*: It is the proportion of semantic, phonemic paraphasia, morphological errors, and neologisms with respect to total number of words.

Appendix D

Analysis of RST parameters

Table D1.
Descriptive Statistics of Four RST Parameters on Refused Umbrella

Parametric		Fluency-1st half	Fluency-2nd half	Size of relation set	Type-token ratio
Aphasic	Mean (SD)	22.61 (13.23)	22.10 (10.19)	8.40 (4.05)	0.64 (0.12)
	Range	7.50-53.99	6.79-42.86	3.00-15.00	0.40-0.82
Control	Mean (SD)	37.96 (9.56)	36.75 (6.80)	10.93 (1.91)	0.59 (0.07)
	Range	21.43-60.01	21.43-50.00	7.00-13.00	0.50-0.72
Non-parametric		No. of EDUs	No. of relations	Depth of structure	% Structural expansion
Aphasic	Mean (SD)	14.60 (5.18)	11.47 (5.80)	5.27 (1.79)	1.77 (3.93)
	Range	7.00-26.00	4.00-24.00	2.00-8.00	0.00-12.50
Control	Mean (SD)	19.20 (4.71)	17.67 (4.47)	6.80 (1.42)	0.67 (1.76)
	Range	10.00-26.00	9.00-24.00	5.00-10.00	0.00-5.00
		% Incomplete EDU	% Failed EDUs	% Functional words	% Errors
Aphasic	Mean (SD)	5.38 (10.32)	6.47 (19.25)	48.31 (7.99)	4.87 (4.44)
	Range	0.00-36.36	0.00-75.00	34.04-58.67	0.00-15.38
Control	Mean	1.59 (3.26)	0.00	52.27 (4.76)	0.21 (0.56)
	Range	0.00-11.76	0.00	47.46-64.29	0.00-1.69

Table D2.
Descriptive Statistics of Four RST Parameters on Egg and Ham Sandwich

Parametric		Fluency-1st half	Fluency-2nd half	Size of relation set	Type-token ratio
Aphasic	Mean (SD)	16.22 (7.97)	21.42 (8.24)	3.60 (1.45)	0.72 (0.12)
	Range	4.61-30.00	4.29-35.00	1.00-6-00	0.55-0.90
Control	Mean (SD)	23.28 (6.43)	28.41 (7.73)	5.53 (3.20)	0.61 (0.10)
	Range	15.00-36.01	16.88-50.00	2.00-11.00	0.45-0.85
Non-parametric		No. of EDUs	No. of relations	Depth of structure	% Structural expansion
Aphasic	Mean (SD)	8.40 (4.05)	4.80 (2.24)	2.80 (1.01)	3.02 (6.49)
	Range	4.00-19.00	1.00-8.00	1.00-4.00	0.00-20.00
Control	Mean (SD)	14.00 (6.31)	8.80 (4.65)	4.80 (2.48)	1.06 (2.22)
	Range	7.00-23.00	3.00-16.00	2.00-9.00	0.00-6.25
		% Incomplete EDU	% Failed EDUs	% Functional words	% Errors
Aphasic	Mean (SD)	11.32 (14.20)	4.79 (10.31)	50.22 (8.47)	6.72 (5.79)
	Range	0.00-50.00	0.00-33.33	32.00-64.44	0.00-18.18
Control	Mean (SD)	2.85 (5.05)	0.00	53.84 (6.43)	0.17 (0.65)
	Range	0.00-14.29	0.00	40.74-64.71	0.00-2.50

Table D3
Descriptive Statistics of the Results of Perceptual Judgment by Naïve Listeners

Refused Umbrella		Coherence rating	% Completeness	% Order	% Unclarity
Aphasic	Mean (SD)	4.60 (1.60)	42.67 (28.21)	86.67 (19.52)	65.00 (29.28)
	Range	1.40-6.70	5.00-95.00	25.00-100.00	10.00-100.00
Control	Mean (SD)	7.95 (0.54)	90.95 (14.19)	98.67 (2.97)	9.42 (10.17)
	Range	7.15-9.00	50.00-100.00	90.00-100.00	0.00-30.00
Sandwich		Coherence rating	% Completeness	% Order	% Unclarity
Aphasic	Mean (SD)	4.71 (1.74)	50.68 (34.87)	81.56 (14.45)	54.96 (28.61)
	Range	1.95-6.55	0.00-90.00	60.00-100.00	5.00-100.00
Control	Mean (SD)	7.33 (0.93)	79.21 (21.99)	94.91 (7.34)	15.21 (16.25)
	Range	6.25-9.00	25.00-100.00	73.68 -100.00	0.00-60.00

Appendix E

Distribution of relation types

Fig E1. *Line Graph of the Distribution of Relations of Refused Umbrella*

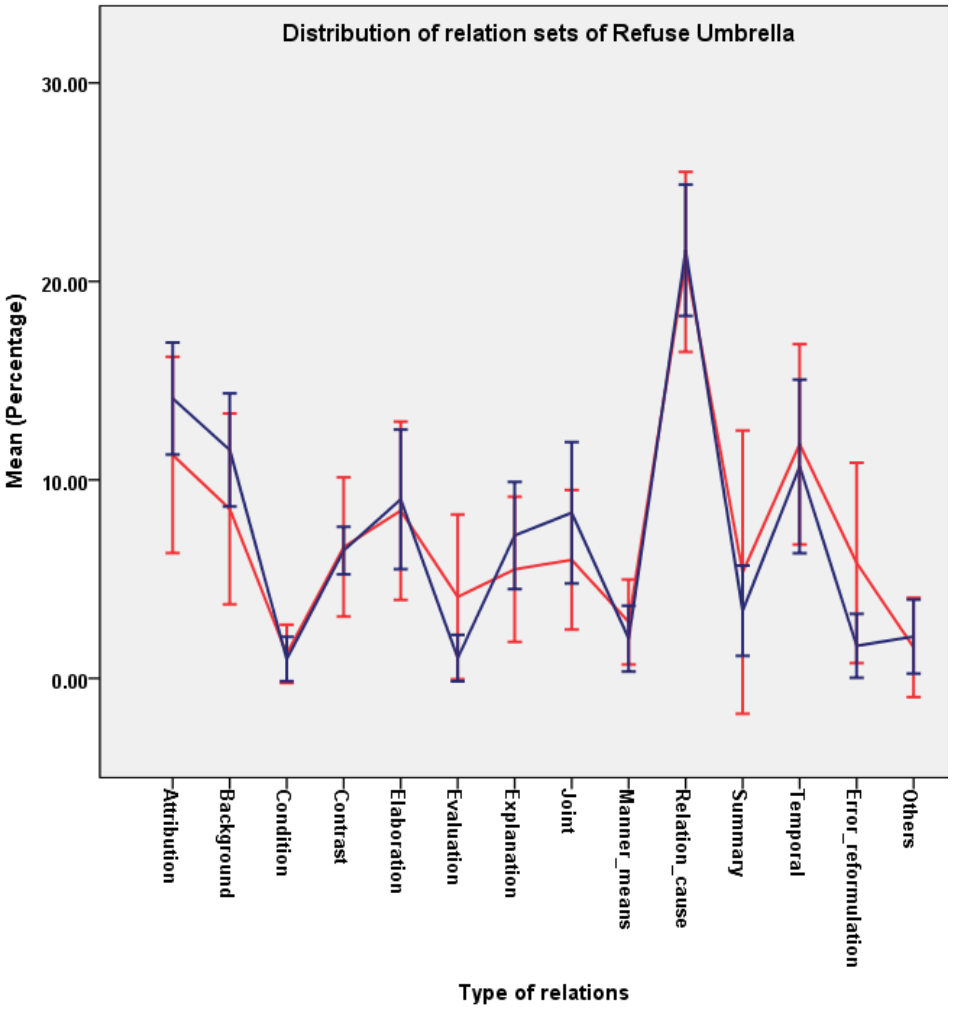
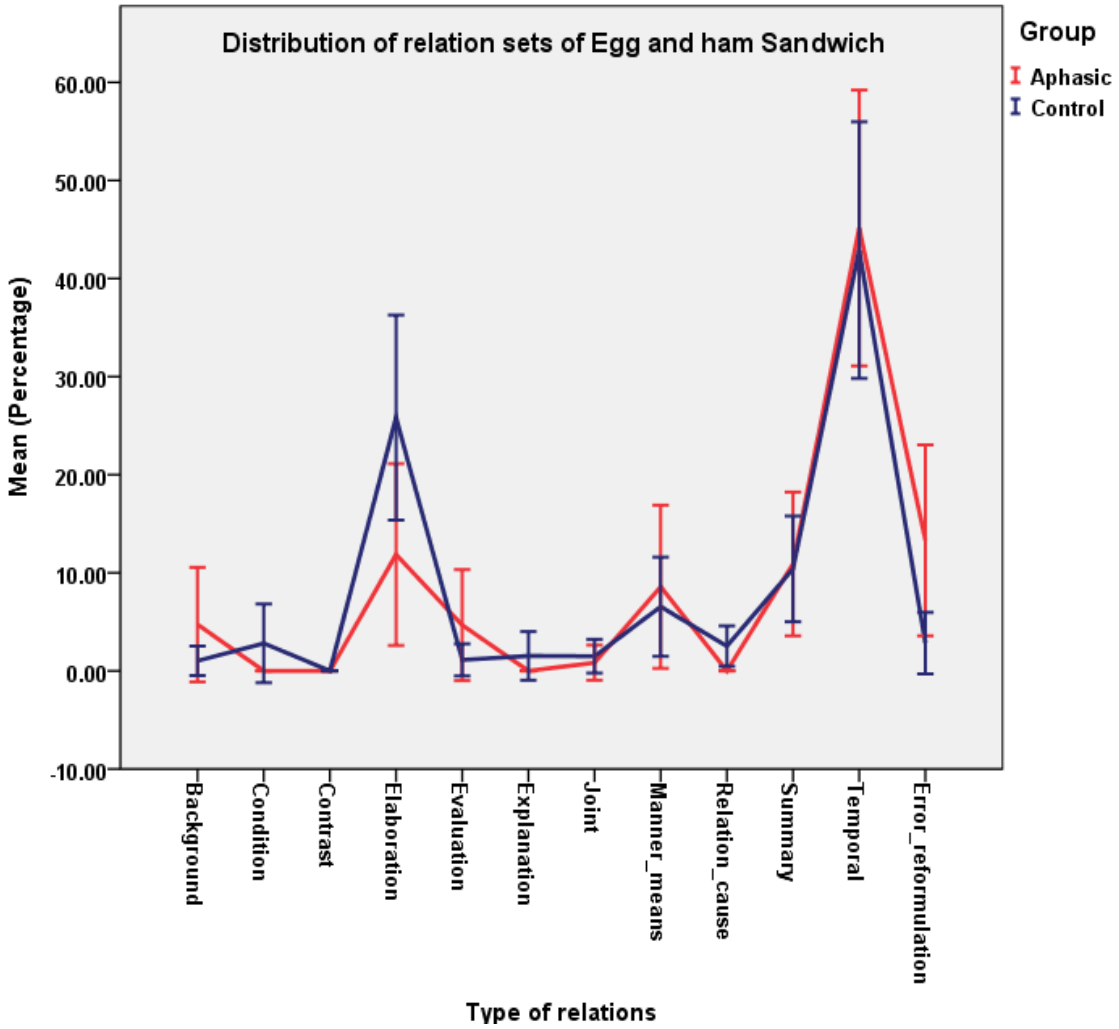


Fig E2. *Line Graph of the Distribution of Relations of Sandwich*



Appendix F

Correlation between RST parameters and subjective measurements

Table 1

Correlations between RST Parameters with Subjective Measurements

Refused Umbrella

Measures	Fluency_1st half	Fluency_2nd half	No. of EDU	Size of relation set	No. of relation	Depth of structure	Structural expansion	Incomplete EDU	Failed EDU	TTR	Functor	Error
Coherence score	**.425*** **	**.544*** **	**.474*** **	**.560*** **	**.570*** **	**.472*** **	-.017(ns)	-.032(ns)	**-.389** **	-.206 (ns)	.188 (ns)	**-.709 *** **
Complete-ness	**.408** **	**.489*** **	**.482*** **	**.512*** **	**.583*** **	**-.454*** **	.013(ns)	.017(ns)	**-.362***	**-.277***	.096 (ns)	**-.728*** **
Order	**.301***	**.384** **	**.291***	**.373***	**.418** **	**-.360***	.066(ns)	.031(ns)	**-.486** **	-.020 (ns)	.053 (ns)	**-.456** **
Unclarity	**-.433*** **	**-.484*** **	**-.307***	**-.395** **	**-.386** **	**-.301***	-.077(ns)	.161(ns)	**.403** **	.185 (ns)	-.059 (ns)	**.563*** **

Egg and ham Sandwich

Measures	Fluency_1st half	Fluency_2nd half	No. of EDU	Size of relation set	No. of relation	Depth of structure	Structural expansion	Incomplete EDU	Failed EDU	TTR	Functor	Error
Coherence score	**.263***	.161(ns)	**.362** **	.248(ns)	**.344** **	**.431** **	-.023(ns)	**-.449** **	**-.369***	**-.256***	.241(ns)	**-.623*** **
Complete-ness	.104(ns)	.212(ns)	**.306***	**.271***	**.297***	**.361** **	-.109(ns)	**-.293***	**-.303***	-.153 (ns)	.161(ns)	**-.455** **
Order	.131(ns)	.151(ns)	.202 (ns)	.127(ns)	.169(ns)	.213(ns)	-.062(ns)	**-.403** **	**-.318***	-.050 (ns)	.121(ns)	**-.382***
Unclarity	-.196(ns)	-.147(ns)	-.125 (ns)	-.088(ns)	-.137(ns)	**-.289***	.101(ns)	**.516*** **	**.403** **	.078 (ns)	-.133(ns)	**.555*** **

Note: ns: non-significant;*p < .05; **p ≤ .01; *** p ≤ .001

Appendix G

Components in narrative discourse

The classification of essential and elaborative components (Labov, 1972):

a. *Orientation*: It gives background and setting information.
b. *Complicating action*: It tells the next event in response to a potential question "And what happened [then]?".
c. *Evaluation*: It is the assessment or emotional comment on the consequences.
d. *Resolution*: It finishes off the event and resolves any complication
e. *Coda*: It closes the story and connects the ending to the present context.

Essential components in narrative discourse are components that carry the critical informational load (orientation, complicating action, and resolution). For procedural discourse, essential components are the steps necessary to complete the task. In our story-telling task, the shortest sample comprising both the essential and elaborative components was chosen from the control group as the exemplar because it made up the simplest form of a complete narrative structure. The shortest sample with essential procedural steps was extracted from the control's sequential description. The agreement of the presence of essential components between the aphasic production and the exemplar in both tasks was computed and presented below.

Table 1
Agreement of Essential Components between Groups

Refused Umbrella			
Orientation	a. Mum gave the kid an umbrella (93%)	**Complicating action**	c. The child went out (60%)
	b. The child refused (93%)		d. It started to rain (100%)
Evaluation (Optional)	g. Mum was angry (13%)		e. The child went home (33%)
	h. Mum gave the child the umbrella (20%)		f. The child got drenched (53%)
Resolution	i. The child brought the umbrella(100%)	**Coda (Optional)**	k. That's the end of the story (0%)
	j. The child went out (33%)		
Egg & Ham sandwich			
Procedures	a. Beat an egg (Optional) (40%)	b. Whisk the egg (Optional) (27%)	c. Pan-fry the egg (93%)
	d. Pan-fry the ham (93%)	e. Put on a piece of bread (100%)	